IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Insomnia And Sleep Apnea Disease Detection Using Machine Learning

Dr.G.SriLakshmi[1], Ch.Priyanka[2], Y.Veera Mahesh[2], K.Shashank sheshagiri [2] , A.Sowmika[2]

Associate Professor[1], Department of information technology and engineering, SRK Institute of Technology, NTR, Andhra Pradesh, India

Student[2], Department of information technology and engineering, SRK Institute of Technology, NTR, Andhra Pradesh, India

ABSTRACT

Sleep disorders can significantly impact an individual's overall health and well-being, making early detection crucial for effective intervention. This project presents a machine learning-based approach to detecting and classifying sleep disorders using health-related data. The dataset used includes key features such as sleep duration, quality of sleep, physical activity levels, stress levels, body mass index (BMI), blood pressure, heart rate, and daily step count. To ensure accurate classification, multiple preprocessing steps are applied, including categorical encoding, feature scaling, and handling missing or duplicate values. The system utilizes a Random Forest classifier, trained on preprocessed data to predict potential sleep disorders. The model is evaluated using accuracy metrics and a classification report to ensure its reliability. The functionality is made accessible through a Flask-based web application, where users can input their health parameters and receive instant predictions regarding their sleep health. The system is designed to assist individuals in understanding their sleep patterns and taking necessary steps toward improved well-being. By leveraging machine learning, this project provides an efficient and user-friendly tool for sleep disorder prediction and awareness.

Key Words: Sleep Disorders, Machine Learning, Random Forest Classifier, Flask Web Application.

INTRODUCTION

In this project, we explore the application of machine learning techniques to detect and differentiate between insomnia and sleep apnea using key health-related features such as sleep duration, quality of sleep, body mass index (BMI), stress levels, heart rate, blood pressure, and daily activity levels. By leveraging supervised learning algorithms such as Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), the system aims to classify sleep disorders with high accuracy. The integration of ML with a Flask-based web interface provides an accessible tool for users to input their data and receive instant feedback regarding their sleep health. This approach not only enhances accessibility but also supports early intervention and awareness, ultimately contributing to better sleep hygiene and overall well-being.

Sleep is a vital physiological process essential for maintaining physical health, cognitive function, and emotional well-being. However, sleep disorders such as insomnia and sleep apnea can significantly disrupt sleep quality, leading to adverse health outcomes including fatigue, impaired concentration, cardiovascular diseases, and metabolic disorders.

Insomnia is characterized by difficulty falling or staying asleep, while sleep apnea involves repeated interruptions in breathing during sleep, often due to airway obstruction or irregular brain signals. Early and accurate detection of these disorders is critical for timely treatment and improving overall health outcomes.

Traditional diagnostic methods, such as polysomnography and clinical evaluations, are often expensive, time-consuming, and require specialized facilities. To address these challenges, machine learning (ML) offers a promising alternative by enabling automated, data-driven diagnosis based on physiological and behavioral health indicators.

PROBLEM STATEMENT

Traditional methods of diagnosing sleep disorders, such as polysomnography, present significant challenges, including high costs, extensive time commitments, and the necessity for specialized equipment and trained professionals. These barriers result in many individuals suffering from various sleep disorders remaining undiagnosed, as the availability of these diagnostic tools is often limited or entirely inaccessible to them. In light of these issues, there is an urgent need for a more accessible, cost-effective, and efficient approach to identifying and predicting sleep disorders. Leveraging advanced machine learning models that can analyze easily obtainable health and lifestyle data offers a promising solution. Such an approach could empower individuals to gain insights into their sleep health, facilitate early diagnosis, and ultimately contribute to improving the overall quality of life for those affected by sleep disorders. Sleep disorders such as Insomnia and Sleep Apnea are increasingly affecting individuals across various age groups due to factors like stress, sedentary lifestyles, and poor sleep hygiene. These conditions, if left undiagnosed, can lead to serious long-term health issues including cardiovascular diseases, hypertension, and mental health disorders. Traditional diagnostic methods often require costly and time-consuming clinical tests (e.g., polysomnography). Hence, there is a growing need for automated, accessible, and reliable screening tools that can help individuals and healthcare professionals in early detection.

MOTIVATION

This constructs a Flask web application designed to predict a health-related outcome. It leverages pre-trained machine learning models, including popular algorithms like Random Forest and Support Vector Machines, to analyze user-provided health data. The application handles data input through a web form, performs necessary preprocessing steps such as encoding categorical features and scaling numerical ones, and then utilizes the loaded models to generate predictions. By presenting the results through a user-friendly web interface, this code bridges the gap between complex machine learning models and accessible health insights.

Here are five key aspects of this code:

1. Model Integration: The application seamlessly loads and utilizes multiple pre-trained machine learning models (ANN, DecisionTree, KNN, RandomForest, SVM), showcasing a capability for model comparison or future ensemble methods to potentially enhance prediction accuracy.
2. Data Preprocessing Pipeline: It implements a clear and structured data preprocessing pipeline, including loading pre-fitted label encoders and scalers, and a safe_encode function to handle potentially unseen categorical values during prediction, ensuring data consistency and model compatibility.
3. Flask Web Interface: The use of Flask creates a straightforward web interface with a homepage (/) for user interaction and a dedicated prediction endpoint (/predict) that receives user input via a form and displays the prediction results on a separate page (result.html).
4. Organized Structure: The code is well-organized with clear sections for global variables, model loading, safe encoding, and Flask routes, improving readability, maintainability, and allowing for easier expansion or modification of the application's functionality.
5. Testing Capability: The inclusion of a test_prediction function demonstrates a proactive approach to verifying the functionality of the prediction endpoint through a programmatic HTTP request, which is crucial for development and ensuring the API operates correctly.

LITERATURE REVIEW

1. Talal Sahreed., Alshammari.,(2024) Applying Machine Learning Algorithms for the Classification of Sleep Disorders

This paper investigates the use of machine learning algorithms, specifically KNN, SVM, Decision Tree, and Random Forest, to classify sleep disorders using the Sleep Health and Lifestyle Dataset. By analyzing daily activities and lifestyle habits, the study aims to develop non-invasive methods for early diagnosis of conditions like insomnia and apnea. The research highlights the superior accuracy of SVM (92.04%) and Random Forest (91.15%) in handling the multidimensional health data for effective sleep disorder detection. Ultimately, this work contributes to data-driven techniques for identifying potential sleep-related health risks through comparative model analysis.

2. Malathi.D., Selvapriya.S.,(2025) Machine learning approaches for sleep disorder prediction

This paper by Malathi and Selvapriya .This research explores machine learning approaches for predicting sleep disorders, utilizing algorithms such as k-nearest neighbors, support vector machine, decision tree, and random forest. The study is conducted by Malathi.D and Selvapriya.S and employs the Sleep Health and Lifestyle Dataset, focusing on features related to sleep patterns and daily activities. The evaluation of various models reveals promising accuracy rates, with Gradient Boosting Classifier achieving the highest at 95%, followed by ANN at 92.92% and SVM at 92.04% on the given dataset size. This work highlights the potential of machine learning in effectively identifying and potentially predicting sleep-related health issues based on relevant lifestyle data.3.Volpe, M., & Das, J. (2023). methylR: a graphical interface for comprehensive DNA methylation array data analysis. Bioinformatics.

3. Puneet KumarYadav.,Uday Kumar Singh.,(2023) Sleep Disorder Detection Using Machine Learning Method

This paper by Puneet Kumar Yadav, Uday Kumar Singh, Judeson Antony, and Tamilarasi R. (2023) explores the use of Decision Tree, Logistic Regression, and XGBoost algorithms. The dataset used includes features such as gender, BMI level, and sleep disorder types. Among the models tested, the Decision Tree achieved the highest accuracy of 93.5%, followed by Logistic Regression with 92% and XGBoost with 72%. The study highlights model performance but lacks clarity on clinical application.

4. Shivam Tivari.,Deepak Arora.,(2019)Supervised Approach Based Sleep Disorder Detection Using Dynamic Features Of EEG

Taran et al. (2019) described how sleep apnea is a sleeping problem that negatively impacts health and the use of EEG readings to identify apnea occurrences. Li et al. (2019) observed that people have alternating interhemispheric activity during sleep, and sleep apnea exacerbates this imbalance. Tiwari et al. explained that autism spectrum disorder (ASD) is a complicated and diverse neurodevelopmental disease and also discussed that abnormal brain connection is the primary neurobiological process behind it. Taran et al. (2021) suggested that electroencephalogram (EEG) signals with a single characteristic Lampel-Ziv complexity can be used to identify sleep apnea.

5. Limo.,(2024) Research on Sleep Health Prediction and Algorithms Based on Big Data

Shanbhog and Medikonda (2023) explore the relationships between sleep quality using student data. Sadeghi, Banerjee, and Hughes (2020) propose that electronic health records and heart rate variability can predict sleep quality. Dvir H. et al. (2020) analyze central sleep apnea (CSA) in comparison to healthy and obstructive sleep apnea (OSA) patients. Biddle, Hermens, Lallukka, Aji, and Glozier (2019) suggest that insomnia and short sleep are linked to heart health issues.

6. Almodovar Cruz, G. E., Kaunitz, G., Stein, J. E., Sander, I., Hollmann, T., Cottrell, T. R., ... & Sunshine, J. C. (2022). Immune cell subsets in interface cutaneous immune- related adverse events associated with anti- PD- 1 therapy resemble acute graft versus host disease more than lichen planus. [6]

Almodovar Cruz et al. (2022) conducted a study on immune cell subsets in interface cutaneous immune-related adverse events (irAEs) associated with anti-PD-1 therapy. They observed that these immune cell subsets have similarities to acute graft-versushost disease (GVHD) rather than lichen planus. This research could be significant for advancing our understanding of immune responses in irAEs and improving the management and treatment strategies for patients receiving anti-PD-1 therapy.

7. Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). More than you've asked for: A Comprehensive Analysis of Novel Prompt Injection Threats to Application-Integrated Large Language Models. [7]

The study conducted by Greshake et al. (2023) explores the novel prompt injection threats to application-integrated large language models. The authors provide a comprehensive analysis, going beyond what was initially asked for, to uncover potential vulnerabilities and security risks. The research sheds light on the importance of understanding these threats for the advancement and improvement of large language models integrated into various applications. The findings of this analysis could have significant implications for the development and deployment of such models.

8. Van Doorn, S., Heyden, M. L., Reimer, M., Buyl, T., & Volberda, H. W. (2022). Internal and external interfaces of the executive suite: Advancing research on the porous bounds of strategic leadership.

The article by Van Doorn et al. (2022) explores the internal and external interfaces of the executive suite and how they contribute to strategic leadership. It highlights the porous bounds of strategic leadership and the need for research to advance our understanding in this area. The authors argue that a comprehensive career advancement interface is vital for executives to navigate and succeed in complex 9 business environments. This study adds valuable insights to the field of strategic leadership and provides a foundation for further research in this domain.

9. Jung, Y., & Sohn, Y. W. (2022). Does work passion benefit or hinder employee's career commitment? The mediating role of work–family interface and the moderating role of autonomy support. [9]

The study conducted by Jung and Sohn (2022) explores the relationship between work passion and career commitment. They examine the mediating role of work-family interface and the moderating role of autonomy support in this relationship, providing insights into how these factors influence comprehensive career advancement.

10. Ammar, A. (2022). Distance Learning and on Job Training from a Messenger Pigeon to Implanted Brain Microchips Interface and Metaverse. In Learning and Career Development in Neurosurgery: Values-Based Medical Education. Cham: Springer International Publishing. [10]

In his book, "Learning and Career Development in Neurosurgery: Values-Based Medical Education," Ammar (2022) explores the concept of distance learning and onthe-job training, from traditional methods such as messenger pigeons to cutting-edge technologies like implanted brain microchips interface and the metaverse. This comprehensive approach to career advancement highlights the importance of integrating new technologies into the field of neurosurgery and emphasizes the value of lifelong learning for medical professionals.

EXISTING SYSTEM:

The existing system is a Flask-based machine learning web application designed to predict sleep disorders such as insomnia and sleep apnea using health-related user input. Users interact with the system through a simple web interface where they enter parameters like age, gender, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI category, blood pressure, heart rate, and daily step count. Upon submission, the backend processes the data by applying label encoding to categorical variables and scaling numerical features using pre-trained encoders and a scaler. The system loads multiple machine learning models—Random Forest, SVM, ANN, Decision Tree, and KNN—trained offline and saved as .pkl files. Each

model predicts the sleep disorder class based on the processed input, and the predictions are then decoded into readable labels using a target encoder. The final prediction, primarily from the Random Forest model, is displayed on the results page. The application also includes a test function that simulates a POST request to verify model functionality. Although the system is functional and user-friendly, it currently relies on manual data entry, lacks integration with real-time data sources like wearable devices, and does not store historical data or provide personalized recommendations.

PROPOSED SYSTEM:

In the proposed system, we introduce a web-based machine learning application designed to predict an individual's health-related condition based on lifestyle and biometric parameters. Built using the Flask framework, this application integrates multiple pre-trained machine learning models to deliver accurate and robust predictions.The system processes user-submitted data, including variables such as age, gender, occupation, sleep duration, physical activity, stress level, BMI category, blood pressure, heart rate, and daily steps. To ensure consistency and reliability, categorical inputs are encoded using label encoders, and numerical features are standardized using a trained scaler. The input data is then evaluated by several machine learning models including Artificial Neural Networks (ANN), Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest.Each model generates a prediction, which is translated from encoded form into human-readable labels using a target encoder. Among these, the system selects the prediction from the Random Forest model as the final output, due to its high reliability and performance in classification tasks.The application includes a user-friendly interface for input and result display, making it accessible for non-technical users. Additionally, the system is modular, allowing for the easy integration of additional models or replacement of existing components.This solution aims to support users in understanding their health status by offering predictive insights based on everyday health metrics, thus empowering proactive health management through accessible technology.

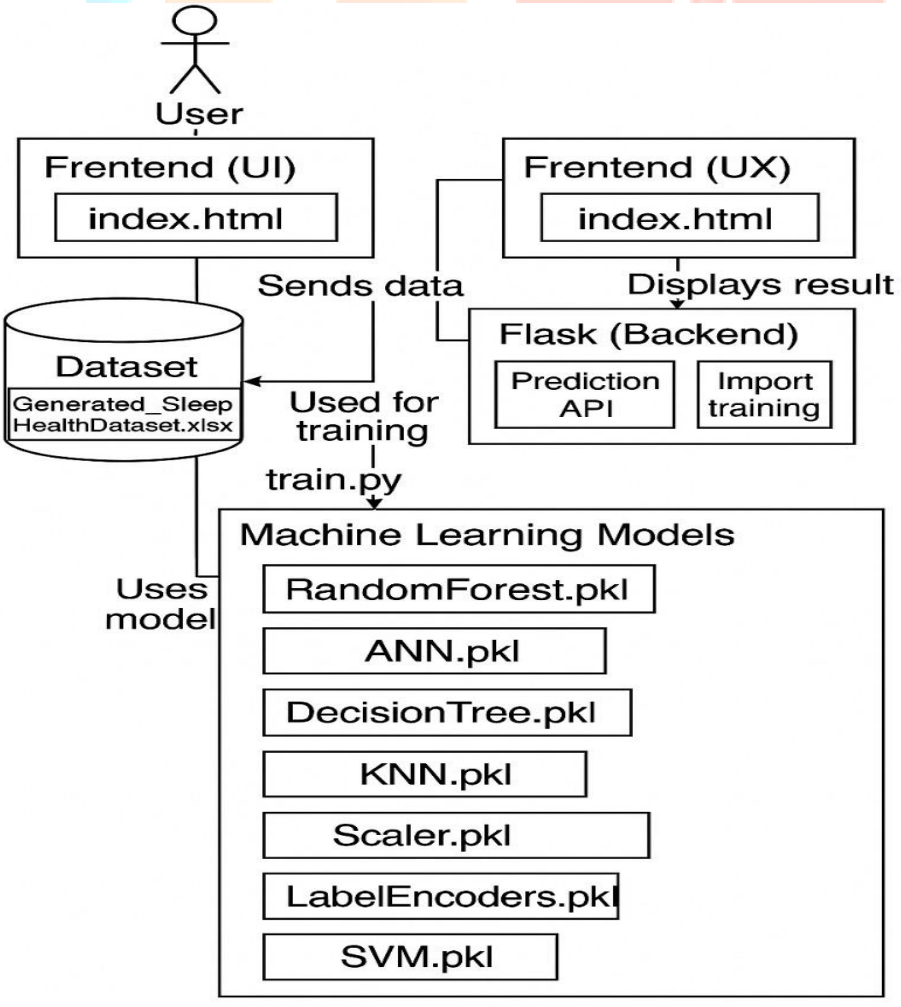

Figure 1: Block Diagram for Proposed System

METHADOLOGY: Sleep Disorder Prediction System

1.Requirement Analysis

- A simple interface for inputting health and lifestyle data, Quick and accurate prediction of sleep disorder types, Minimal clinical dependency

Real-time response and high model interpretability

- System responsiveness (<2 seconds for prediction), User engagement and feedback satisfaction

2. System Design

➢ Architecture Overview

• Frontend: HTML templates (index.html, result.html)
• Backend: Flask framework in app.py
• ML Models: Trained using Scikit-learn (train.py)
• Storage: Models, encoders, and scalers in model_files/

➢ Database/Storage

• Excel-based dataset (Generated_Sleep_Health_Dataset.xlsx) for training
• Encoded, scaled, and stored components using Pickle for runtime loading

3. AI-Powered Career Recommendation System

➢ Machine Learning Models Used

• Random Forest (default model for predictions)
• Artificial Neural Network (MLPClassifier)
• Support Vector Machine (SVC)
• K-Nearest Neighbors
• Decision Tree

➢ Techniques and Libraries

• Label Encoding for categorical features
• Standard Scaling for numerical inputs
• Scikit-learn for model training, evaluation, and saving

➢ Target Outcome

Prediction of three sleep states: Insomnia, Sleep Apnea, or None

➢ Accuracy Monitoring

Models evaluated using classification_report and accuracy_score

4. Skill Gap Analysis and Learning Integration

- Analyze correlation of features like “Stress Level” and “Sleep Duration” with target
- Remove unnecessary columns (e.g., Person ID)
- Handle missing or malformed 'Blood Pressure' by splitting into systolic/diastolic

Remove duplicates for improved training stability

- Dynamic data intake (e.g., from wearables or APIs)

User-specific learning insights on how to improve sleep

5. Intelligent Job Matching and Application Assistance

- Users input daily lifestyle and physiological metrics into a form on the frontend
- LabelEncoders for categorical values, StandardScaler for numerical values
- Predictions from all models logge, Random Forest’s output selected for display
- Real-time error feedback, Safe encoding for unknown input values.

6. Networking and Mentorship Features

- Connect users to sleep coaches or therapists via APIs, Integrate with health portals or wellness networks
- Allow users to anonymously share sleep patterns or concerns, Discussion boards with health experts
- Push reminders about consistent sleep hygiene, Webinars on mental health and sleep management

7. System Implementation

- Flask (Python web framework)

• Scikit-learn for model training and prediction
• Pickle for saving models and encoders

- HTML templates rendered with Flask (render_template)

• Responsive design for input forms and results
- app.py: Main Flask server
• train.py: Model training and saving logic
• model_files/: Stores models and encoders

8. Testing and Validation

- Train-test split with 80:20 ratio, Evaluation using classification metrics (precision, recall, F1-score)
- Test Flask API with real and mock inputs (test_prediction() in app.py)
- Simulate end-to-end form submission, Observe behavior under correct and incorrect inputs
- Safe encoding to handle new/unseen values, Try/Except blocks to handle runtime exceptions

9. Deployment and Maintenance

- Local deployment on Flask server, Ready for deployment on Heroku, AWS, or GCP
- Console logs for predictions and system behavior, Model-wise prediction outputs for transparency
- Modular loading of models allows for quick retraining, Can integrate newer models or updated datasets easily

RESULTS & ANALYSIS

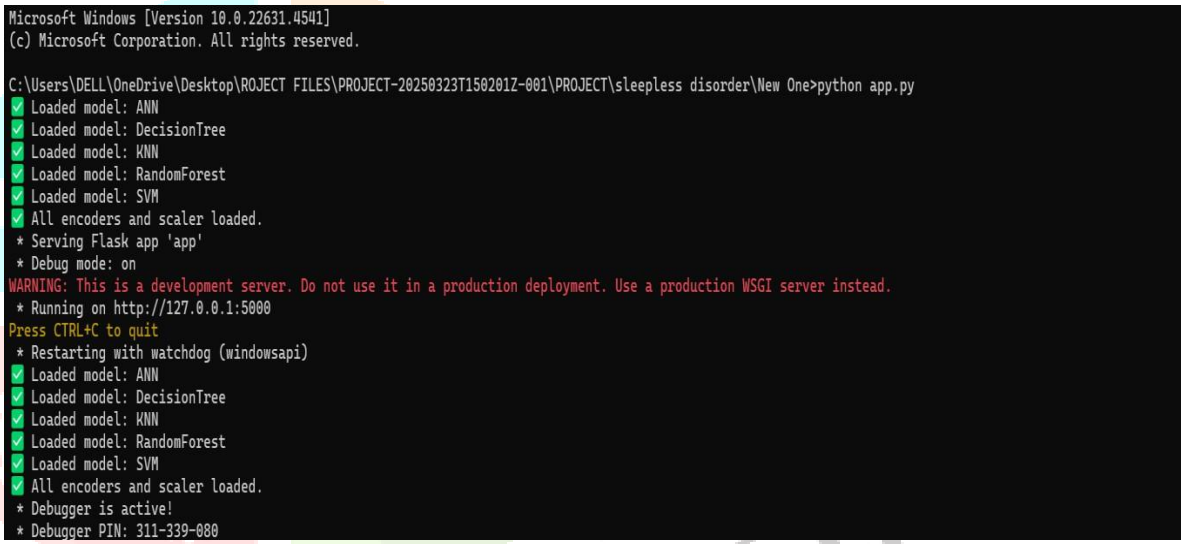

Figure1: To run Python web server

Figure2: Website redirecting

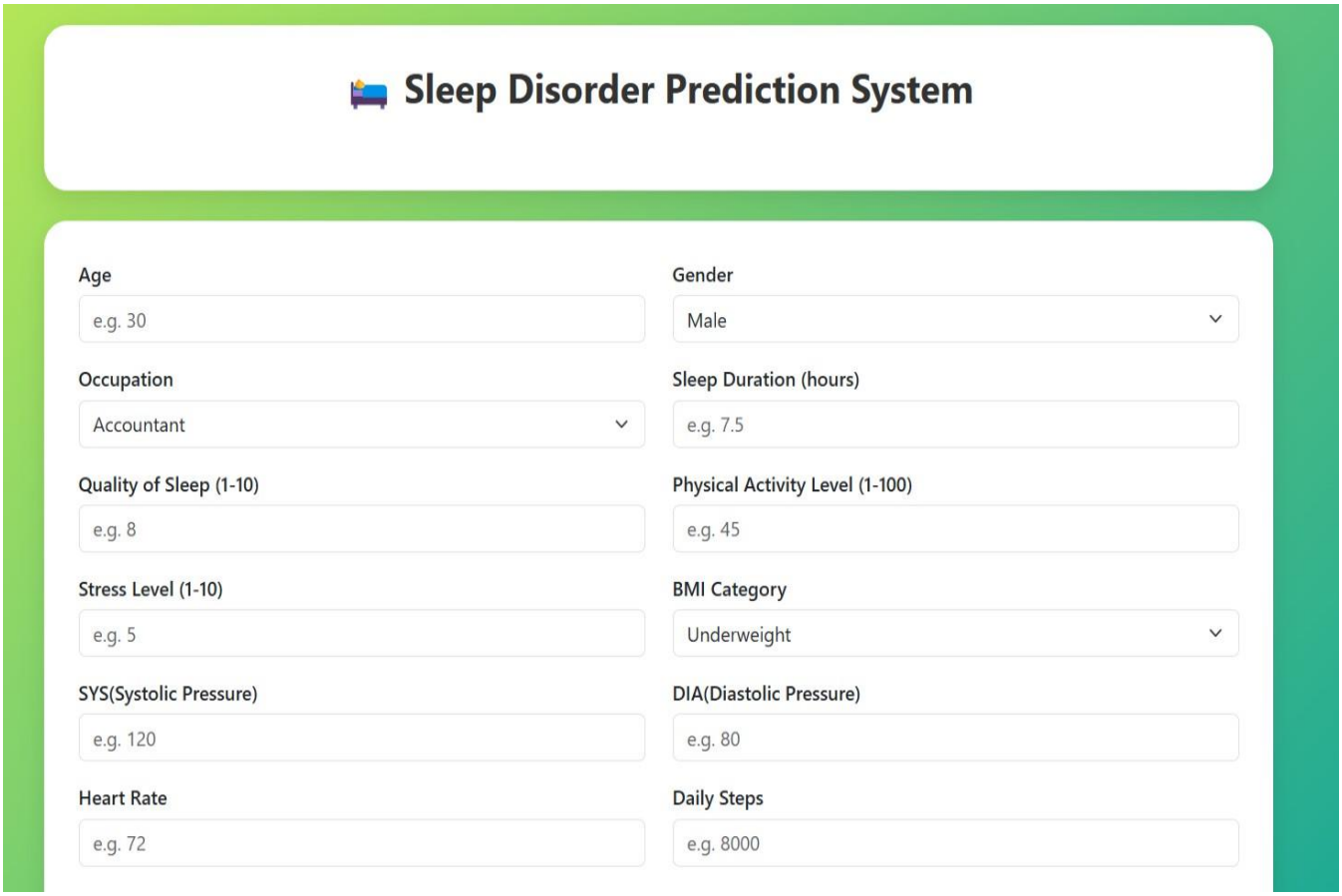

Figure3: Home Page to fill in the details

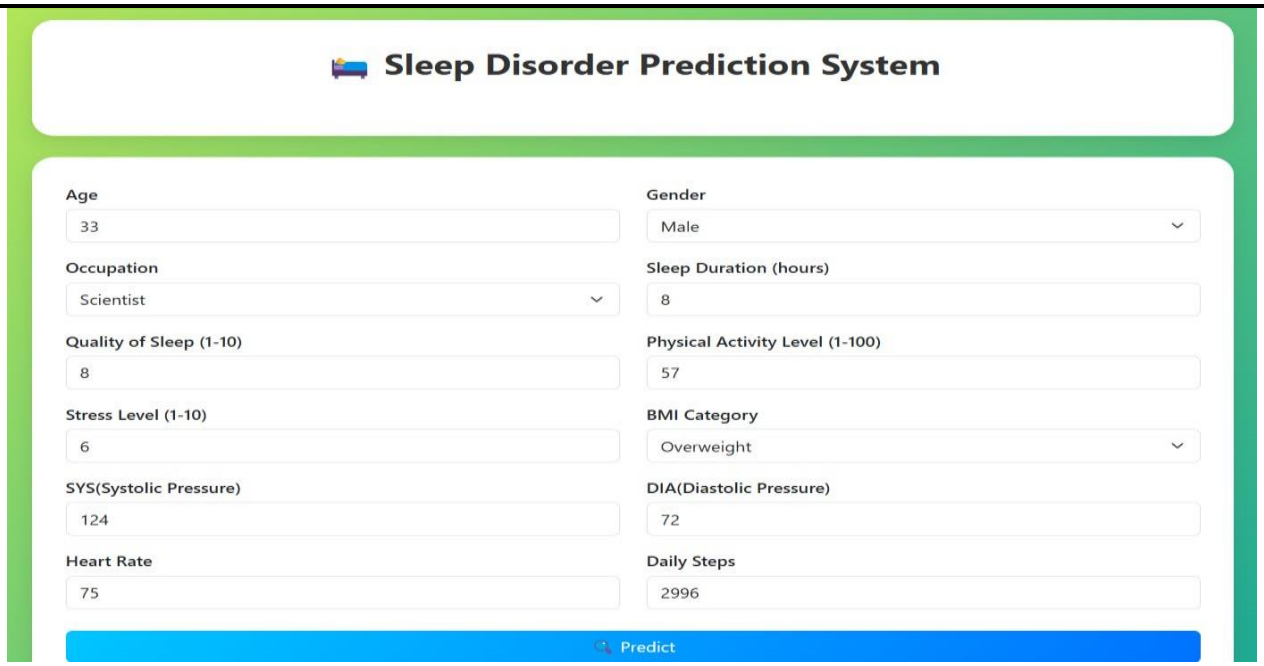

Figure4: Predicting the Output

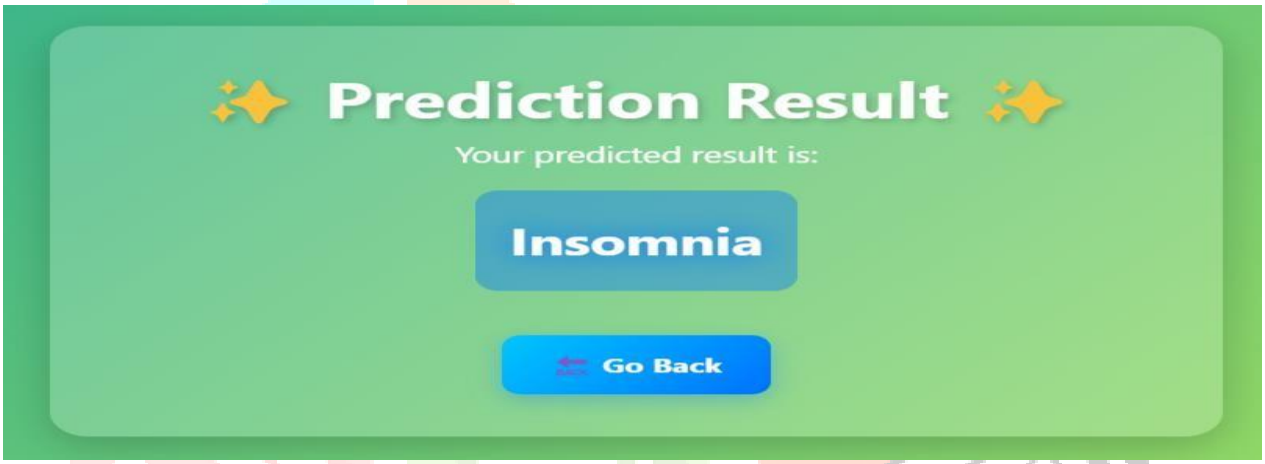

Figure5: Output Window

CONCLUSION

The Sleep Disorder Prediction System is a user-centric, AI-driven platform designed to simplify the early detection of sleep disorders using health and lifestyle inputs. It integrates multiple machine learning models to analyze key health parameters such as stress level, blood pressure, heart rate, and sleep duration, providing reliable predictions for sleep conditions like insomnia and sleep apnea. The system is scalable, modular, and accessible, built using a Flask-based web interface supported by a robust backend architecture for preprocessing and prediction. It addresses the limitations of traditional diagnostic methods by offering an automated, non-invasive, and cost-effective approach. With features like real-time inference, safe encoding, and model performance tracking, the platform serves as a practical tool for both public use and research. Future enhancements include wearable integration, user feedback loops, and visual analytics, ensuring the system evolves with emerging needs in digital health diagnostics.

REFRENCES

[1] Limbu, N., Alsadoon, A., Prasad, P.W.C., et al. (2022). A novel solution of deep learning for sleep apnea detection: enhancement of SC and elimination of GVICS. Multimedia Tools and Applications,81,38569–38592.
https://doi.org/10.1007/s11042-022-13142-1

[2] Ramachandran, A., & Karuppiah, A. (2021). A Survey on Recent Advances in Machine Learning Based Sleep Apnea Detection Systems. Healthcare, 9(7), 914. https://doi.org/10.3390/healthcare907091

[3] Nguyen, A.T., Nguyen, T., Le, H.K., Pham, H.H., & Do, C. (2022). A novel deep learning- based approach for sleep apnea detection using single-lead ECG signals. arXiv preprint arXiv:2208.03408. https://arxiv.org/abs/2208.03408

[4] Borah, S., Gogoi, P., Gohain, P., Boro, C., & Muchahari, M.K. (2022). Machine Learning for Detection of Obstructive Sleep Apnoea. In: Satapathy, S.C., Bhateja, V., Favorskaya, M.N., Adilakshmi, T. (eds) Smart Intelligent Computing and Applications, Volume 2. Smart Innovation, Systems and Technologies, vol 283. Springer, Singapore. https://doi.org/10.1007/978-981-16-9705-0_24

[5] Mangal, A., & Uttam, A.K. (2020). Sleep Prediction by Various Supervised Machine Learning Models. International Journal of Advanced Science and Technology, 29(05), 3786– 3792.
https://sersc.org/journals/index.php/IJAST/article/view/12288

[6] Mousavi, S., Afghah, F., & Acharya, U.R. (2019). SleepEEGNet: Automated Sleep Stage Scoring with Sequence to Sequence Deep Learning Approach. arXiv preprint arXiv:1903.02108. https://arxiv.org/abs/1903.02108

[7] National Sleep Foundation. (2021). Sleep Health Index. Retrieved from https://www.sleepfoundation.org

[8] World Health Organization. (2020). Healthy sleep and chronic disease risk. WHO Technical Report.

[9] Mullen, A. B., & Guo, C. (2023). Data-driven insights into lifestyle impacts on sleep quality: An AI approach. IEEE Access.

[10] Choudhury, T., & Singh, D. (2022). Sleep disorder diagnosis using AI models: A machine learning perspective. International Journal of Health Informatics.